The Report Committee for Adam Jared Torres
Certifies that this is the approved version of the following report:

Wastewater Expenditure Effects on In-stream Bacteria Pollution in the Rio Grande / Río Bravo Post-NAFTA: Evidence from Panel Data Estimations

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Chandler Stolp

Co-Supervisor: ______________________________
Sheila Olmstead

Wastewater Expenditure Effects on In-stream Bacteria Pollution in the Rio Grande / Río Bravo Post-NAFTA: Evidence from Panel Data Estimations

by

Adam Jared Torres, B.S.; B.A

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Public Affairs

And

Master of Arts

The University of Texas at Austin

August 2014

Dedication

First and foremost, this report is dedicated to my parents Ned and Ruthanne Torres. Thank you for providing constant support and encouragement throughout my time in graduate school and helping me overcome the many obstacles involved with the process. I would also like to dedicate this work to my supportive wife, Caitlin, who has been there to lift me up all along the way. Without her patience, encouragement, and hope, I would not have had the resolve to finish. I dedicate this report to my brother Dustin, who has always believed in me. Lastly, this paper is dedicated in the loving memory of my former graduate advisor Shama Gamkhar.

Acknowledgements

This report would not have been possible without the guidance and supervision of Sheila Olmstead, Ph.D. and Chandler Stolp, Ph.D. Special thanks to the Texas Commission on Environmental Quality and the Lozano Long Foundation for the financial support over the years for my work on this research. At TCEQ, I would like to thank Roger Miranda for his continuous support and encouragement throughout this process, Steve Niemeyer for sharing ideas and knowledge, and Richard Hay for providing insight with the use of NEXRAD data. At the LBJ School of Public Affairs, I would like to thank David Eaton Ph.D. for the sharing of knowledge and Shama Gamkhar Ph.D. and Pat Wong Ph.D. for their steadfast advising. At the Lozano Long Institute of Latin American Studies, I am very grateful for the academic advising of Lorraine Leu Ph.D. and the constant support of Steve Alvarez. Thank you to my friends and colleagues Ross Van Horn and Rachel Daggy for sharing ideas and inspiration during the writing of this report. A very special thank you goes out to my wife, Caitlin, who has remained a pillar of support throughout this journey.

Abstract

Wastewater Expenditure Effects on In-stream Bacteria Pollution in the Rio Grande / Río Bravo Post-NAFTA: Evidence from Panel Data Estimations

Adam Jared Torres, MPAff.; MA

The University of Texas at Austin, 2014

Supervisors: Chandler Stolp, Sheila Olmstead

The United States and Mexico share responsibility in preserving the quality of their international river system, the Rio Grande / Río Bravo, and several international treaties govern the quantity of water each country must give and take. Because no treaty establishes joint standards for the quality of the river, the North American Agreement on Environmental Cooperation (NAAEC) was created in 1993 as a declaration of principles and objectives concerning the conservation and the protection of the environment as well as a guide of concrete measures to further cooperate on these matters. One particular goal of the NAAEC was to improve water quality in the US-Mexico Border Region, ensuring a clean, safe, and reliable water supply for the area. Although the US and Mexican federal governments have made substantial technical and financial commitments through binational agencies like the North American Development Bank (NADB) and the Border Environment Cooperation Commission (BECC), few empirical

studies have assessed the impact of binational expenditures on wastewater infrastructure in this region. This report uses longitudinal panel data regression models to estimate the impact of capital expenditures on water quality made by binational, federal, and state water quality management institutions from 1995 to 2012. This analysis considers expenditures made on both sides of the Rio Grande watershed that constitutes the international border, beginning with El Paso, Texas and ending in the Gulf of Mexico.

Table of Contents

List of Tables

List of Figures

List of Acronyms

BECC	Border Environment Cooperation Commission
BEHI	Border Environment Health Initiative
BEIF	Border Environment Infrastructure Fund
CAFO	Concentrated Animal Feeding Operations
CAP	Community Assistance Program
CFU	Colony-Forming Units
CILA	Comisión Internacional de Límites de Aguas
CONAGUA	Comisión Nacional de Agua
CRP	Clean Rivers Program
CRWR	Center for Research in Water Resources
CWSRF	Clean Water State Revolving Fund
CWTAP	Colonias Wastewater Treatment Assistance Program
DEM	Digital Elevation Model
DF	Degrees of Freedom
DWSRF	Drinking Water State Revolving Fund
EDAP	Economically Distressed Areas Program
EPA	United States Environmental Protection Agency
FCS	Fully Conditional Specification
FE	Fixed Effects
GIS	Geographic Information Systems
HUC	Hydraulic Unit Code
IBWC	International Boundary and Water Commission
INEGI	Instituto Nacional de Estadística y Geografía

MI	Multiple Imputation
MPN	Most Probable Number
NAAEC	North American Agreement on Environmental Cooperation
NADB	North American Development Bank
NAFTA	North American Free Trade Agreement
NAIP	National Agriculture Imagery Program
NEXRAD	Next-Generation Radar
NHD	National Hydrology Dataset
NOAA	National Oceanic and Atmospheric Administration
NSDI	National Spatial Data Infrastructure
PDAP	Project Development Assistance Program
RE	Random Effects
SPS	Sanitary and Phytosanitary Measures
SRM	Standards Related Measures
SWQMIS	Surface Water Quality Monitoring Information System
TCEQ	Texas Commission on Environmental Quality
TMDL	Total Maximum Daily Load
TNRIS	Texas Natural Resources Information System
TWDB	Texas Water Development Board
USGS	United States Geological Survey
WDF	Water Development Fund
WWTP	Waste Water Treatment Plant

CHAPTER 1: INTRODUCTION

Background

The Rio Grande – or Río Bravo, as the river is called in Mexico – is the fifth longest river in North America and among the 20 longest in the world. It originates in the San Juan Mountains of Colorado, runs North-South the length of New Mexico, and turns east in El Paso (U.S.) and Ciudad Juarez (Mexico) to form a 1,254-mile national border with Mexico before emptying into the Gulf of Mexico at Brownsville (U.S.) and Matamoros (Mexico). The entire Río Bravo watershed covers an area of approximately 924,300 square kilometers (335,000 square miles), with approximately half the watershed in the United States and the other half in Mexico. The Texas Commission on Environmental Quality (TCEQ) has divided the river into 14 segments based on varying hydrologic conditions. This report will address segments 2314, 2307, 2306, 2304, and 2302, spanning three discrete sections of the river – the Upper Rio Grande, the Middle Rio Grande, and the Lower Rio Grande.

The Need for Environmental Infrastructure

Whether in the form of sewage and solid waste control, access to potable water, or wastewater treatment, environmental infrastructure has long been a focus of both the United States and Mexican governments. For some time, one of the most severe environmental problems facing the border has been the challenge of supplying the region with safe and reliable water. Water is very scarce along the border and resources to finance needed infrastructure have been in even shorter supply. Surface water resources have become and remain seriously threatened by the border region's rapid industrialization coupled with increased population growth. Rapid industrial and population growth in border cities continues to outstrip their environmental infrastructure, including that for water supply, wastewater collection and treatment, municipal solid waste disposal, and industrial hazardous waste disposal. Thousands of residents on both sides of the border lack access to wastewater collection, and untreated wastewater is a major transboundary externality, as raw or partially treated sewage flows into drinking water sources on both sides of the border. In 1995, only 40% of the border region residents in Mexico had access to potable water (Giner, 2009). Sewage infrastructure coverage in 1995 was estimated to be 35-50% for Mexican Border States, and wastewater treatment coverage was minimal (Giner, 2009).

Over 20 years ago, the United States and Mexico took part in groundbreaking negotiations to strengthen their economic ties, resulting in the signing of the North American Free Trade Agreement (NAFTA) (North American Free Trade Agreement, 1992). As the first international trade agreement to include environmental provisions, an important legacy of NAFTA was the adherence to the North American Agreement on Environmental Cooperation (NAAEC), which set out to address the environmental impacts of trade liberalization in North America (NAAEC, 1993). As result of the passage of the NAAEC, the Border Environment Cooperation Commission (BECC) and North American Development Bank (NADB) were created to mitigate possible environmental impacts of rapid population and industrial growth in the region. Since the BECC has been in existence, 234 border infrastructure projects have been certified for funding and the NADB has financed 171 of those projects with the support of US$1.9 billion in loans and grants that are benefiting an estimated 17.2 million residents (NADB, 2012). The idea behind the BECC and NADB is to strengthen and continue binational reciprocity through the construction and management of environmental and human health infrastructure. Despite the investment of considerable resources over the past 20 years, there has yet to be a comprehensive assessment of its long-term effectiveness.

An Overview of the Rio Grande Basin

The portion of the Rio Grande / Río Bravo that forms an international boundary between the United States and Mexico encompasses an immense area from the arid Chihuahuan Desert region near El Paso, Texas, to the Gulf of Mexico. The Rio Grande Basin spans three ecological regions of Texas (the Trans-Pecos, Edwards Plateau, and South Texas Plains) that are characterized by their similarity of climate, landform, geology, soil, natural vegetation, and hydrology. Because water quality resets at the beginning of a reservoir, it is necessary to divide the Rio Grande Basin into three sub-basins. Each of these river basins exhibit different characteristics in river length, ecosystem types, and precipitation.

The Upper Rio Grande Sub-basin

The Upper Rio Grande Sub-basin extends a length of 650 miles (1,045 km) from the New Mexico-Texas state line downstream to the International Amistad Reservoir (Figure 1, p. 22). Due to geomorphic and historical changes in the channel, the river meanders in and out of Texas and New Mexico with some sections forming the boundary between the two states. Continuing downstream, the Rio Grande forms the international boundary between the U.S. and Mexico. This region's economy relies on agriculture, manufacturing, tourism, wholesale and retail trade, and government operations, including the Fort Bliss Army installation in El Paso, Texas.

The Upper Rio Grande Sub-basin lies entirely in the Trans-Pecos eco-region. For its water supply, this region largely depends on the Edwards-Trinity Plateau and Hueco-Mesilla Bolsons aquifers combined with six minor aquifers. The upper portion of the Rio Grande traverses the mountains of the Chihuahuan desert, wandering through arid mesas, steep hills, and rock outcrops as it passes through Big Bend National Park.

Water in the Upper Rio Grande is used for agriculture by New Mexico, Texas, and Mexico during the irrigation season. The City of El Paso also uses the river to provide half of its drinking water supply (IBWC, 2008). The sister cities of El Paso, Texas, and Ciudad Juarez, Chihuahua, have a combined population of more than 2 million, and lands surrounding the cities are used primarily for agriculture. Downstream, the river is greatly influenced by agricultural return flows, wastewater effluent, and raw or partially treated sewage. Thus, the Upper Rio Grande downstream of El Paso and Ciudad Juarez is characterized by high levels of salinity and bacteria. The Rio Conchos tributary joins the Rio Grande at the sister cities of Presidio, Texas, and Ojinaga, Chihuahua, improving water quality and significantly increasing water quantity (IBWC, 2008).

The Middle Rio Grande Sub-basin

The Middle Rio Grande Sub-basin contains the segment of the Rio Grande flowing just below International Amistad Reservoir and extends its reach to the International Falcón Reservoir. This portion of the river stretches 303 miles (487 km) and traverses five counties in Texas as well as the Mexican states of Coahuila, Nuevo Leon, and Tamaulipas. The majority of the population living along the Rio Grande in this reach dwell in the cities of Del Rio, Eagle Pass and Laredo, Texas along with Mexican sister cities Ciudad Acuña, Coahuila, Piedras Negras, Coahula, and Nuevo Laredo, Tamaulipas.

The easternmost and northernmost portions of the Middle Rio Grande Sub-basin lie in the Edwards Plateau eco-region while the remaining Sub-basin falls under the Texas Brush Country eco-region. In the Middle Rio Grande Sub-basin downstream of the International Amistad Reservoir, the terrain transitions to create irregular and rolling plains until it approaches the coastal plains closer to the Lower Rio Grande Sub-basin. Most cities and municipalities along this portion of the river depend on surface water for domestic, agricultural, and industrial use. This region's economy relies primarily on tourism, hunting, ranching, and government operations (IBWC, 2008).

The Lower Rio Grande Sub-basin

The Lower Rio Grande Sub-basin spans 280 miles (451 km) from just below the International Falcón Dam to its confluence with the Gulf of Mexico. In Texas, this part of the river acts as a border between Starr, Hidalgo, and Cameron counties and the Mexican state of Tamaulipas. Population has rapidly increased in the Lower Rio Grande sub-basin in the past ten years (IBWC, 2008). This region's economy is heavily dependent on agriculture, as well as trade, services, manufacturing, and hydrocarbon production. Major cities and municipalities in this sub-basin include McAllen, Harlingen, and Brownsville, Texas, and Matamoros and Reynosa, Tamaulipas. This region depends wholly on the Rio Grande/Río Bravo for drinking water. Rapidly increasing population and industrialization add further constraints to this region, which routinely experiences drought conditions and high agricultural water use.

The southeastern portion of the South Texas Brush Country eco-region is occupied by the Lower Rio Grande Sub-basin. The two major aquifers for this region are the Carrizo-Wilcox and Gulf Coast aquifers and groundwater in this region is brackish, presenting future desalinization plant construction possibilities for the region. Most agricultural and urban discharges do not enter the Rio Grande in this Sub-basin, because they are diverted to canals that ultimately empty into the Gulf of Mexico.

Summary of Water Quality Issues

The Texas Commission on Environmental Quality (TCEQ) characterizes the Rio Grande (Río Bravo) by either classified or unclassified segments (IBWC, 2008). Classified segments, also referred to as designated segments, are water bodies that are protected by site-specific criteria. Classified water bodies include major rivers and their tributaries, major reservoirs, and estuaries. Unclassified waters are typically smaller water bodies without site-specific water quality standards assigned to them but are protected instead by general surface water quality standards that apply to all surface water in the state of Texas (IBWC, 2008).

This study contains five classified segments of the Rio Grande. If the data assessed by TCEQ indicates poor water quality, the water body may receive a classification of "impaired" because it is not supporting its designated use. The water quality impairments identified by the TCEQ in its latest assessment cycle of water quality testing for all five segments in this study are shown in Table 1. All segments, except for segment 2306 are currently classified as "impaired" for *E. coli* bacteria concentration.

Table 1: Bacteria Impairments in the Rio Grande Watershed

River Section	Segment	Segment Name	Bacteria Impairment	Year First Listed
Upper Rio Grande	2314	Above International Dam	*E. coli*	2002
	2307	Below Riverside Diversion Dam	*E. coli*	2002
Middle Rio Grande	2306	Above International Amistad Reservoir	-	-
	2304	Below International Amistad Reservoir	*E. coli*	1996
Lower Rio Grande	2302	Below International Falcon Reservoir	*E. coli*	1996

This report is divided into five chapters. Chapter two will review current and past literature surrounding in-stream bacteria concentration in the Rio Grande, followed by a historical analysis of what transboundary water quality management institutions have done to invest in environmental infrastructure improvements along the border. Chapter three will cover the methodology and data acquisition of this project, describing the sources of data and the implications of each variable's descriptive statistics. Chapter four presents the analytical methods involved in this study and is divided into five sub-sections: addressing missing data in this study, using multiple imputation to estimate data missingness, using a panel data linear regression approach for program evaluation, deciding fixed versus random effects in the linear regression model, and model specification based on variables included in this study. Lastly, chapter five presents the results of the series of fixed effects panel data regressions, providing key insights into the effects of capital expenditures, if any, on in-stream bacteria in each of the three sections of the Rio Grande. At the end of chapter five, insights are provided for the recommendation of future study.

CHAPTER 2: LITERATURE REVIEW

In-stream Bacterial Pollution and Human Health

Health Risk from Infected Waters

Fecal coliform data have historically been used to indicate potential risk of illness that may occur from exposure to polluted waters for contact recreation. When fecal pollution exists in a water body used for contact recreation, individuals are most commonly at risk to illness-causing organisms through accidental ingestion of contaminated water. Fecal coliform does not cause illness directly, but is a good indicator of any number of harmful pathogens that commonly co-exist in water bodies where they are present.

Human pathogens represent a subset of the immeasurable microorganisms present in the environment at any given point of time. Human pathogens cause varying degrees of illness (Environmental Protection Agency, 2000; Environmental Protection Agency, 2001). Some of these pathogens are naturally-occurring in vegetation, but the majority are present in the feces or other wastes of humans and other warm-blooded animals. Gastroenteritis, the most frequent result of infection by these pathogens, is a general term for diseases affecting the gastrointestinal tract and is rarely life-threatening (Environmental Protection Agency, 2002). Symptoms of gastroenteritis may include vomiting, nausea, diarrhea, stomachache, headache, and fever (Environmental Protection Agency, 2001; Environmental Protection Agency, 2002). Persons suffering from gastroenteritis have flu-like symptoms several days after exposure, do not usually suspect that the cause of their illness could be from the ingestion of water, and often assume other causes of their illness (Perez-Ciordia et al., 2002; National Research Council (US) Committee on Indicators for Waterborne Pathogens, 2004). As a result, outbreaks of disease are only intermittently detected and reported, which problematizes determining how many illnesses result from contact with recreational waters.

Bacterial infection from fecal coliform in polluted waters can include cholera, salmonellosis, shigellosis, and gastroenteritis. Viral infections, intestinal diseases caused by enteroviruses, and protozoan infections, such as cryptosporidiosis, amoebic dysentery, and giardiasis, pose dangers associated with fecal coliform as well.

Processes by Which Fecal Coliform Enters a Water Body

Fecal coliform may enter a body of water in a number of ways. The two most common ways fecal coliform enters a water body are by the release of untreated or poorly treated sewage and from precipitation runoff that mixes and drains untreated sewage into a basin. Generally, sources of pollution are classified into point and non-point sources. Pathogenic organisms, such as fecal coliform, are among the many types of pollutants generated at a source (point or non-point) and then transported to a body of water by storm water runoff, groundwater, or other methods. Waste Water Treatment Plants (WWTP) are the most easily identifiable point-sources for fecal coliform in the Rio Grande. During the process of wastewater treatment, there are many factors that may influence the concentration of bacteria that is eventually discharged into a water body. If local WWTPs exhibit inadequate and poorly designed treatment processes, one can expect a spike in the bacterial pollution in the body of water. Other point sources that contribute bacteria into a water body include concentrated animal feeding operations (CAFO), slaughterhouses, meat processing facilities, textiles, pulp and paper facilities, and also fish and shellfish processing facilities. These sources contribute bacterial pollution to a lesser extent than WWTPs, but maintain a substantial load of pathogens and fecal indicators transferred to water bodies (Environmental Protection Agency, 2001).

Non-point sources of pollution are sources of pollution that do not originate from a specific outfall location. The Environmental Protection Agency (EPA) differentiates non-point sources of pollution from point sources, because the former are typically driven by precipitation and other wet-weather events (Environmental Protection Agency, 2001). In documenting the Total Maximum Daily Load (TMDL) protocol for microorganisms in water bodies, the EPA further disaggregates non-point source pollution based on urban or rural sources (Environmental Protection Agency, 2001). Most rural non-point source pollution accumulates from domestic livestock grazing or from wildlife indigenous to an area. Urban non-point sources of bacterial pollution originate from human-generated effects – litter, contaminated refuse, domestic pet and wildlife excrement, and failing sewer lines (Environmental Protection Agency, 2001). Untreated runoff from raw sewage also directly contributes to the overall fecal coliform load in a body of water. Major storm water runoff can greatly increase the bacteria load in a water body and may also contribute to increased organic matter in a water body from the overflow or dysfunction of septic tanks.

Regulatory Standards for Bacterial Pollution in Bodies of Water

Fecal coliform data have historically been used to indicate potential risk of illness that may occur from exposure to polluted waters for contact recreational activities. After 2002, the EPA elected *E. coli* as an indicator species of fecal contamination based on research that has shown *E. coli* to be a better predictor than fecal coliform alone (Texas Clean Rivers Program 2003). *E. coli*, a subset of fecal coliform, is found in the gut of warm-blooded animals, and rarely exists in nature with no connection to fecal matter (Stuart, McFeters, Schillinger & Stuart 1976). Indicators such as fecal coliform and *E. coli* give a measurement of the amount of in-stream fecal coliform present that could be associated with pathogens (Texas Clean Rivers Program 2003).

In the United States, ambient water quality criteria for bacteria were established by the EPA in order to develop a set of guidelines to preserve water quality, standardize procedures involving the assessment of water quality, and to protect citizens from illness due to the exposure to recreational waters (Environmental Protection Agency, 1986). The EPA recommends an *E. coli* geometric mean of no more than 126 colony-forming units (CFU) per 100mL of water and a sample maximum of 394 CFU/100 mL for freshwaters that will be used for primary contact, such as recreation. Primary contact recreational activities include swimming, water skiing, or other activities involving prolonged contact with the water, with considerable risk of ingesting enough water to pose a health hazard (Environmental Protection Agency, 2000). For secondary contact, such as non-contact recreational use of freshwaters, the EPA recommends a geometric mean of 605 CFU/100 mL (Environmental Protection Agency, 2000). Secondary contact recreation are any waters designated for fishing, wading, and boating activities, where contact with the water may occur and there is low probability of ingesting appreciable quantities of water (EPA, 2000).

Mexican water quality standards for bacteria are determined by the permissible limits in the *Secretaría de Salud's NORMA Oficial Mexicana* (NOM 1996). According to these standards, the Mexican drinking water permissible standard limit for total coliform is 2 most probable number (MPN) per 100mL. For fecal coliform, no detectable amount is permissible per 100 mL of water.

Overview of Transboundary Water Quality Management

From a public health perspective, the most pressing water quality issues along the border region are the lack of access to clean drinking water and the lack of sewage treatment. Water contamination resulting from solid waste, raw sewage, and untreated wastewater is suspected to be a key contributing factor to the presence of certain diseases in border populations. Recent research findings have shown alarming water borne disease incidence rates in border counties, with the presence of Hepatitis A, legionella, salmonella, and shigellosis respectively occurring more than three, five, four, and six times higher than the U.S. average (United States Department of Agriculture, Environmental Protection Agency, 2014).

History of Border Water Cooperation

Environmental problems in the Border Region have deep historical roots and cannot be attributed to a single cause but rather to a combination of factors related to fast urban and population growth during the last five decades, combined with the rapid industrialization since the 1970s. Urban growth and industrialization are the result of complex social processes at the local, national, and transnational levels. Binational attention to the region's water resources started in the mid-1800s because a number of rivers define and cross the international boundary. The United States and Mexican governments first formally attempted to address border sanitation problems through the International Boundary and Water Commission (IBWC). A 1944 binational treaty established the IBWC to manage all international water projects and water resource disputes involving the two countries' shared border, including disputes over territorial limits and water allocation. Since the 1970s, rapid industrialization and population growth in the border region created problems that were beyond IBWC's original mandate and resources and led the development of multilateral institutions dedicated toward transboundary water quality management (Table 2).

Table 2: History of Border Water Cooperation

Institution and Date	Description
The International Boundary and Water Commission (IBWC) 1944	The principal binational agency with authority over territorial limits, water allocation, wastewater treatment, sanitation, and water quality.
The La Paz Agreement 1983	Established a framework for cooperation on specific environmental pollution problems. Formal workgroups comprised of federally appointed governmental and academic experts target their policy recommendations toward water, air, contingency planning and emergency response, hazardous waste, enforcement cooperation, and pollution prevention.
North American Free Trade Agreement (NAFTA) 1994	The NAFTA is the first trade agreement which contains provisions to deal with environmental issues which arise in the context of trade relations and disputes: • NAFTA protects certain Multilateral Environmental Agreements from trade challenge (art. 104) • NAFTA prohibits reducing environmental standards to attract investment (arts. 104, 906(2), and 1114). • NAFTA sets general, multilateral rules on Sanitary and Phytosanitary Measures (SPSs) and other Standards Related Measures (SRMs) (arts. 712, 902 and 904). • NAFTA promotes the upward harmonization of environmental policies and standards (arts. 713, 714, 905 and 906). • NAFTA provides for improved consideration of environmental issues in its trade dispute resolution procedures (arts. 723 and 914)
The Border Environmental Cooperation Commission (BECC) 1994	Created to assist border communities and other sponsors in developing and implementing environmental infrastructure projects, and to certify projects for financing consideration by the North American Development Bank or other sources.
The North American Development Bank (NADB) 1994	Capitalized in equal shares by the United States and Mexico to provide $3 billion in new financing to supplement existing sources of funds and leverage the expanded participation of private capital. The BECC/NADB institutions are limited to three types of environmental infrastructure development: water supply and treatment, wastewater treatment and disposal, and municipal solid waste.

Transboundary Water Management Institutions

The International Boundary and Water Commission

The traditional institution for managing transboundary water resources is the IBWC and its Mexican counterpart, the *Comisión Internacional de Límites y Aguas* (CILA). The IBWC headquarters are located in the United States in El Paso, Texas, while CILA main offices operate in Mexico from Juarez, Chihuahua. For more than a century, Mexico and the United States have dealt with transboundary resource issues through this unique binational organization. The IBWC/CILA was founded with the signing of the Convention of 1889 between the two nations in order to resolve differences related to boundary changes of the Rio Grande/ Río Bravo and the Colorado rivers, both of which form segments of the international boundary. Over the next 50 years, the IBWC's primary function was to administer the equitable use and allocation of water in the Rio Grande/Río Bravo.

In 1944, The Treaty on Utilization of Waters of the Colorado and Tijuana and of the Rio Grande further established the IBWC and CILA to oversee and manage all international water projects and water resource disputes involving the two countries' shared border. Henceforward, each organization became responsible to its own national government for:

> "boundary demarcation, channel rectification, construction and maintenance of flood control, water storage, hydroelectric and drainage works, construction and maintenance of sanitation and sewage facilities, scheduling water deliveries under treaty, stream gauging, and the diversion of waters for domestic functions (Mumme, 1993, p. 95)."

Other functions of the IBWC/CILA have been to conduct investigations, execute project planning studies, and adjudicate differences in interpretation of the 1944 Treaty subject to approval of the two governments. According to the 1944 Treaty, the Commission is also authorized to address water sanitation problems by implementing projects mutually agreed upon by the United States and Mexico. These agreements are expressed as “Minutes” of the IBWC/CILA.

The IBWC/CILA is principally a technical agency, addressing water management problems with engineering solutions and scientific evaluations. Although its jurisdiction is limited in scope, it still maintains significant authority. Surrounding issues of U.S-Mexico border water management, the Commission’s authority supersedes claims of other domestic agencies and any attempt to alter its jurisdiction or authority would require a new treaty approved by both governments.

The United States and Mexico signed the Agreement on Cooperation for the Protection and Improvement of the Environment in the Border Area, commonly known as the La Paz Agreement, in 1983. This accord established technical work groups that, for the first time, would address sensitive transboundary issues such as water quality, air quality, natural resources, and solid and hazardous waste. This agreement also established a framework for government agencies of both countries to formally discuss border environmental challenges, exchange data and information, and coordinate plans to reduce pollution in the border area (defined as the area within one hundred kilometers on both sides of the border). In short, the La Paz Agreement has served as an important designator for the intervention of lead federal and state agencies along the border into the arena of border environmental management as it relates to what the IBWC perceives as being beyond its mission and political capabilities.

Border Environmental Cooperation Commission & North American Development Bank

The BECC and the NADB were established in conjunction with NAFTA in 1994. The primary responsibilities of the BECC are to provide technical assistance to border communities and to certify environmental infrastructure projects in the border region for financing consideration by the NADB among other state and federal agencies (BECC/NADB 1999). The NADB's primary purpose is to facilitate financing for the development, execution, and operation of environmental infrastructure projects. Capitalized by both the Mexican and U.S. Governments, the NADB can secure financing at lower commercial rates than border communities could otherwise obtain for commercial loans. The NADB may only finance projects after they have gained certification through the BECC. This division of functions was intended to avoid a conflict of interest: the entity involved in fostering project development (BECC) is different from the organization involved in financing (NADB) (Varady, 1996). The BECC board of directors includes one Administrator each from the USEPA and from *Secretaria del medio ambiente y recursos naturales* (SEMARNAT), Mexico's federal environment ministry. Additionally, the BECC board of directors employs both commissioners from IWBC and CILA, respectively, as well as six additional directors including two state government representatives, two local government representatives, and two public representatives from each country.

The purpose of BECC is "to help preserve, protect, and enhance the environment of the border region in order to advance the well-being of the people in the United States and Mexico" (U.S. Department of State, 1993). By administering a $3 million annual budget appropriated by the U.S. and Mexican legislatures, the BECC promotes and certifies projects that are developed, proposed, and managed by border water and sewer service providers. Certification criteria from the BECC aim to assure investors and border communities that projects meet requirements in the following areas: human health and the environment, technical and financial feasibility, project management, community participation, and sustainable development (see Table 3)

Table 3: BECC Project Certification Criteria

Topical Area	Subject of Certification Criteria
General Criteria	• Project Type • Project Location • Project Description and Work Tasks • Conformance with International Treaties and Agreements
Human Health and Environment	• Human Health and Environmental Need • Environmental Assessment • Compliance with Applicable Environmental and Cultural Resource Laws
Technical Feasibility	• Appropriate Technology • Operation and Maintenance Plan • Compliance with Applicable Design Standards
Financial Feasibility and Project Management	• Financial Feasibility • Fee/Rate Model • Project Management Capacity
Community Participation	• Comprehensive Community Participation Plan • Report Documenting Public Support
Sustainable Development	• Adherence with Sustainable Development Principles • Institutional and Human Capacity Building • Conformance with Applicable Local and Regional Conservation and Development Plans • Natural Resource Conservation

In addition to providing technical assistance, BECC creates and manages coordinating committees engaged in the enhancement of water and sewer services in the border region. These committees typically involve members from the EPA, Comisión Nacional de Agua (CONAGUA, Mexico's Federal water agency), the United States Department of Agriculture, the United States Public Health Service, IBWC/CILA, state utilities of border states from both nations, state governors' offices, state environmental agencies, municipal authorities, as well as local steering committees.

In comparison to the IBWC approach to environmental infrastructure assistance, the BECC certification criteria prioritize a higher level of public participation. Certification criteria require applicants of projects to provide a comprehensive community participation plan, to form a local steering committee, to engage with local environmental organizations, to allow public access to project information, and to hold at least two public meetings during the project's cycle. The service provider responsible for a project (usually a sewer or water utility group) must also submit a report to document the implementation of the community participation plan as well as public support for the project. In addition to external coordination committees, the public also has a role within BECC's organization. A binational Advisory Council, consisting of border residents, also advises the Board of Directors. The Advisory Council is designed to incorporate an avenue for public input into BECC activities and certification and Board of Directors make decisions in meetings open to the public. After the BECC Board of Directors vote to certify a project, it is eligible for a NADB financing package.

Contrary to its name, the North American Development Bank functions differently than traditional development banks such as the World Bank. A traditional development bank transfers capital from wealthier nations to poorer ones but in the case of the NADB, the United States, with an annual gross national product (GNP) of over $6 trillion, capitalized the NADB with the same amount as Mexico, whose GNP was approximately 4% of the United States' GNP in 1994 (Browne, 1996). As of December 2012, the NADB has financed 171 environmentally-beneficial infrastructure projects in the U.S-Mexico border region with a capitalization of $3 billion – $450 million in paid-in capital and $2.55 billion in callable capital (North American Development Bank, 2012). The NADB is authorized to use its paid-in capital to make direct loans to communities and to guarantee payment of a community's non-NADB loans. The Bank's callable capital is money that the U.S. and Mexican federal governments pledged to make available in the unlikely case that a large number of NADB borrowers fail to repay their loans.

Typical financing packages combine NADB loans with loans and grants from other government entities and private investors, adding leverage to NADB's limited funding resources. The NADB provides loans to fill financing gaps that are not covered by the government entities or private investors, and loans that are provided by the NADB are for specific projects, not general programs. The projects must be certified by BECC and be financially self-sustaining;

that is, fees collected for water and sewer services must cover operation and maintenance costs and up to twice the cost of repaying creditors. The capital structure of the NADB allows it to lend to utilities that otherwise have difficulty accessing financing from commercial banks (e.g., the NADB can loan to small utilities borrowing one or two million dollars or less). The NADB also offers other financial services, such as loan guarantees and "gap purchases" of bond issues. In a gap purchase, the NADB buys the portions of a bond offer that are not quickly bought by private investors. The Bank then assists in the financing of projects by acting as an investment bankers, a source of financial advice, and a coordinator of grants and loans from multiple sources.

In 1997, the U.S-Mexico Border Water Infrastructure Program, funded by the US Congress through EPA, began awarding grants to water and wastewater systems in the border region through the Project Development Assistance Program (PDAP) for project development and design and the Border Environment Infrastructure Fund (BEIF) for construction and programs administered by BECC and NADB (North American Development Bank, 1997a). The BEIF provides grants for projects that can be used for construction costs to make a project affordable for a community. For projects that qualify for BEIF assistance, the NADB determines the size of the grant for a project using factors such as the socio-economic characteristics of the area, the water and sewer utility's current debt burden, what other sources of funding may be available, and the project management entity's ability to assume debt. Grants from the fund can be used for border water and wastewater projects that are either in the U.S. or in Mexico, but projects constructed Mexico must demonstrate a benefit the United States. To maintain oversight over the use of BEIF grants, the EPA uses a set of project selection criteria and affordability guidelines and the NADB performs analysis to test a proposed project's eligibility to receive BEIF grants. Projects in the United States that are eligible for BEIF grants adhere to slightly different guidelines than proposed projects in Mexico. Projects in the United States meet assistance eligibility if the project cost per household exceeds 1.7% of the median household income (North American Development Bank, 1997b). On the Mexican side, CONAGUA determines Mexican projects' eligibility for BEIF support by using Mexico's Municipal Poverty Index (North American Development Bank, 1997c).

CHAPTER 3: METHODOLOGY AND DATA ACQUISITION

Watershed Delineation

Sources of pollution are limited geographically to the drainage area of the receiving water body, and the watershed for the Rio Grande / Río Bravo is a representation of the geophysical area that drains into the river (Figure 1, p. 22).

The watershed is divided into basins and sub-basins according to the Hydrologic Unit Code (HUC). Because a watershed ends at a reservoir, the segment of the Rio Grande (Río Bravo) that constitutes the international border must be geospatially delineated into three discrete sections, corresponding to the termination of the Upper Rio Grande at Amistad International Reservoir, the termination of the Middle Rio Grande at Falcón Reservoir, and the termination of the Lower Rio Grande at the mouth of the Rio Grande in the Gulf of Mexico (Figure 2, p. 23). The delineation of these watersheds provides the analysis with a way of defining the land area that contributes the bacteria load in different sections of the river. The watersheds for the U.S. side are taken from the National Hydrography Dataset (NHD) website of the United States Geological Survey (USGS) while the watersheds from Mexico are taken from the Instituto Nacional de Estadística y Geografía (INEGI).

This study uses the geographic location of each of the fourteen water quality stations chosen for analysis to delineate the sub-watersheds that contribute to water quality measured at each station. Using this method, water quality stations serve as drainage locations or "pour points" used to define sub-watersheds of the Larger Rio Grande Río Bravo watershed. Delineation of the sub-watersheds in this study is based on the

Figure 1: Rio Grande / Río Bravo International Boundary Watershed Within Binational Study Area

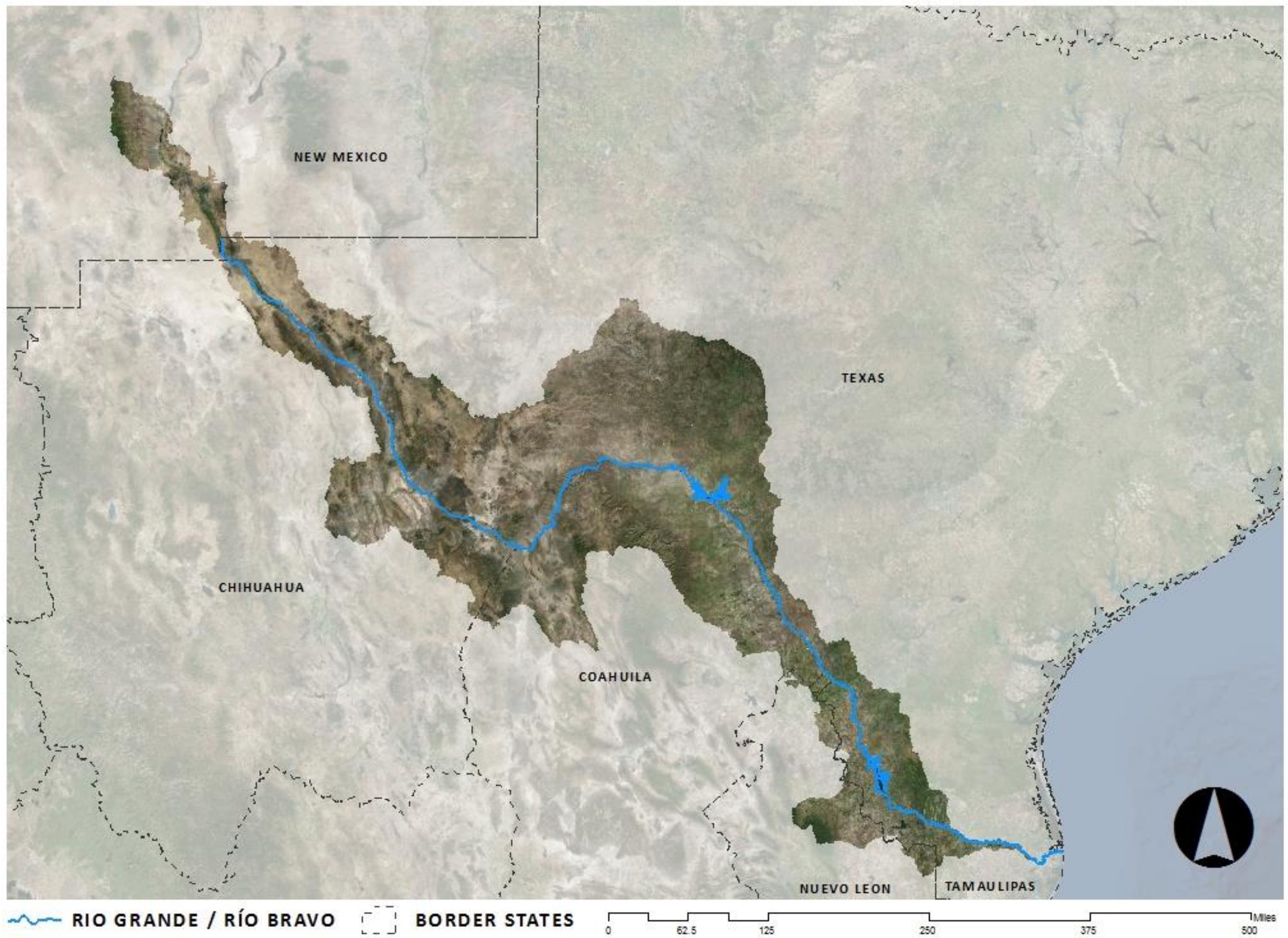
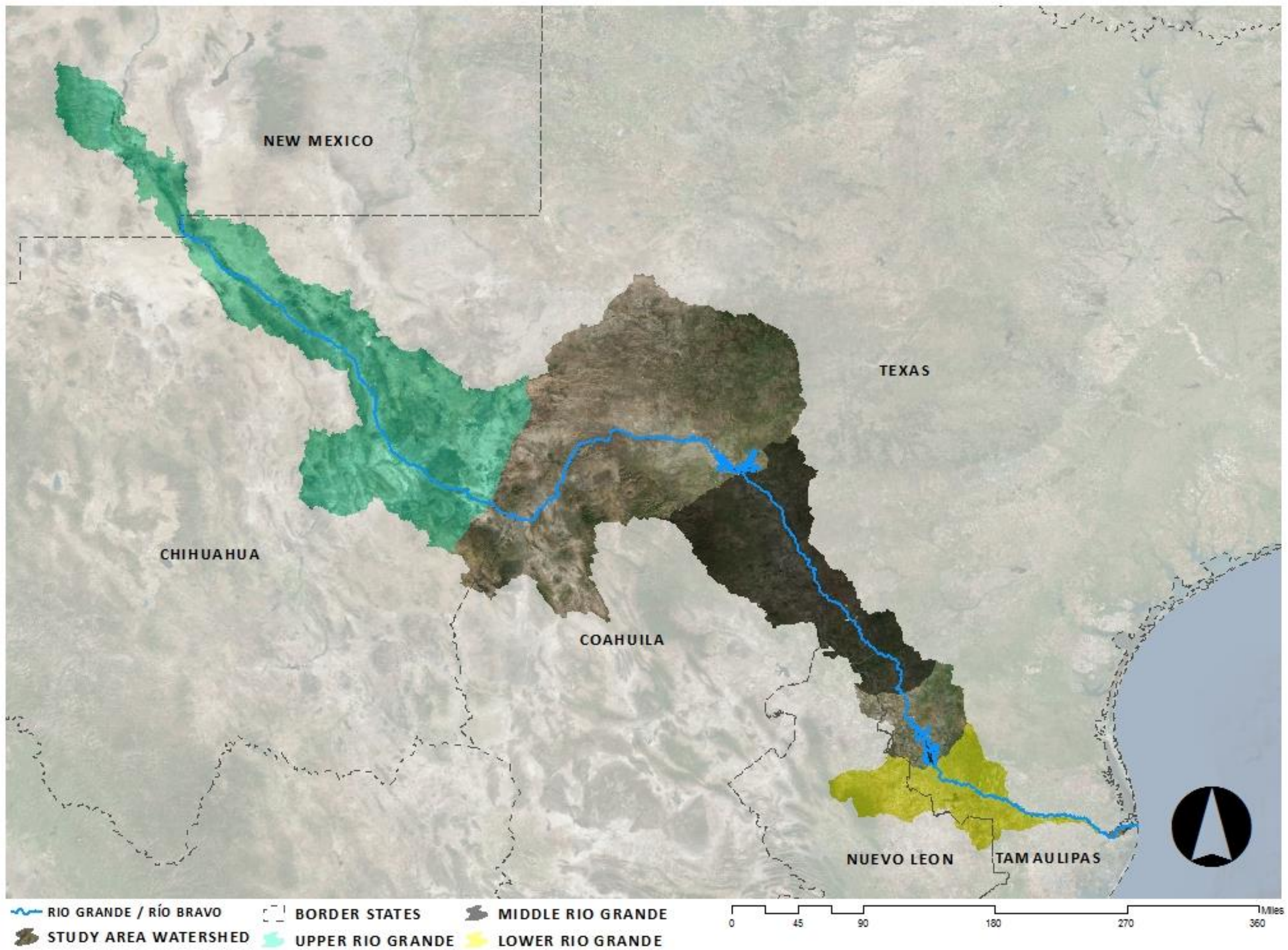


Figure 2: Rio Grande / Río Bravo Basin and Sub-Basins



digital elevation models (DEM) established by the USGS, resulting in watersheds that define geographic areas spatially bound to the study's dependent and independent variables.

Water Quality Data

To assess the risk of disease from waterborne pathogens in fresh surface water and to later test capital expenditures' effectiveness in mitigating that risk, this study confines the dependent variable to *E. coli*, which is measured in the MPN of colony-forming units per one hundred milliliters of water. All historical water quality data from were obtained from sampling stations that are maintained by the partnership between the IBWC's Texas Clean Rivers Program (CRP) and the TCEQ Surface Water Quality Monitoring Information System (SWQMIS). The CRP monitors and assesses the Texas portion of the Rio Grande Basin from the point that it enters the state to its end at the Gulf of Mexico. The CRP conducts chemical, physical, and biological stream surveys and monitoring to assess the quality of receiving streams and to document water quality problem sources and improvements. This study uses *E. coli* data collected at 14 CRP sampling stations that are located at sites above and below municipal or industrial discharges, in areas of significant flow, and that are minimally effected from tributaries, stagnant flow areas, or point sources that could introduce their own chemistry. These stations were chosen based on the length and consistency of the record of bacterial water quality measured at each location as well as the spatial distribution of each station along the Rio Grande / Río Bravo.

In an effort to reflect trends and current conditions, data from SWQMIS as reported by the CRP were included for analysis in this report for the period January 1, 1995 – December 31, 2012. This period of record coincides with the implementation of NAFTA with its corollary NAAEC, and thus corresponds with a period of major infrastructure improvements along the Rio

Grande. Data from the CRP were available through its online database; *E. coli* monitoring events were queried along with water quality parameters associated with this bacteria indicator, including:

- Stream flow (ft^3/sec)
- Water temperature (degrees C)
- Specific conductance (umhos/cm)
- Secchi transparency (m)
- Bio-chemical oxygen demand (BOD – mg/L)
- Ammonia nitrogen (mg/L)
- Nitrate + nitrite nitrogen (mg/L)
- Total phosphorous as P (mg/L)
- Total filterable residue (mg/L)
- Chlorides as CL (mg/L)
- Sulfate as SO4 (mg/L)
- Turbitidy (NTU)

The associated parameters were pulled from the SWQMIS database, averaged on a quarterly basis, and then analyzed using regression and other statistical tests to investigate the relationship between bacteria and other water quality variables, including flow. Another criterion for inclusion in the analysis was the number of samples at each station. Only parameters with ten samples or more were included in regression analysis with *E. coli* data. Generally, stations with less than ten *E. coli* samples were excluded from analysis. Table 4 shows a summary of the TCEQ stations used in the analyses, along with the number of samples available for each station for the entire historical period of record from SWQMIS.

Table 4: Summary of CRP Water Quality Station Data for *E. coli* Samples

Segment	Station	E. Coli Samples	Data Range
2314	13272	155	1995 – 2012
2314	13276	55	1995 – 2012
TOTAL	**2**	**210**	**1995 – 2012**
2307	13230	102	1995 – 2012
TOTAL	**1**	**102**	**1995 – 2012**
2306	13229	100	1995 – 2012
2306	13228	74	1995 – 2012
TOTAL	**2**	**174**	**1995 – 2012**
2304	13208	49	1995 – 2012
2304	13560	98	1995 – 2012
2304	18795	77	1995 – 2012
2304	13202	141	1995 – 2012
2304	15817	104	1995 – 2012
TOTAL	**5**	**469**	**1995 – 2012**
2302	13186	86	1995 – 2012
2302	13184	21	1995 – 2012
2302	13181	71	1995 – 2012
2302	13177	69	1995 – 2012
TOTAL	**4**	**247**	**1995 – 2012**

Location and Description of Water Quality Monitors in the Upper Rio Grande

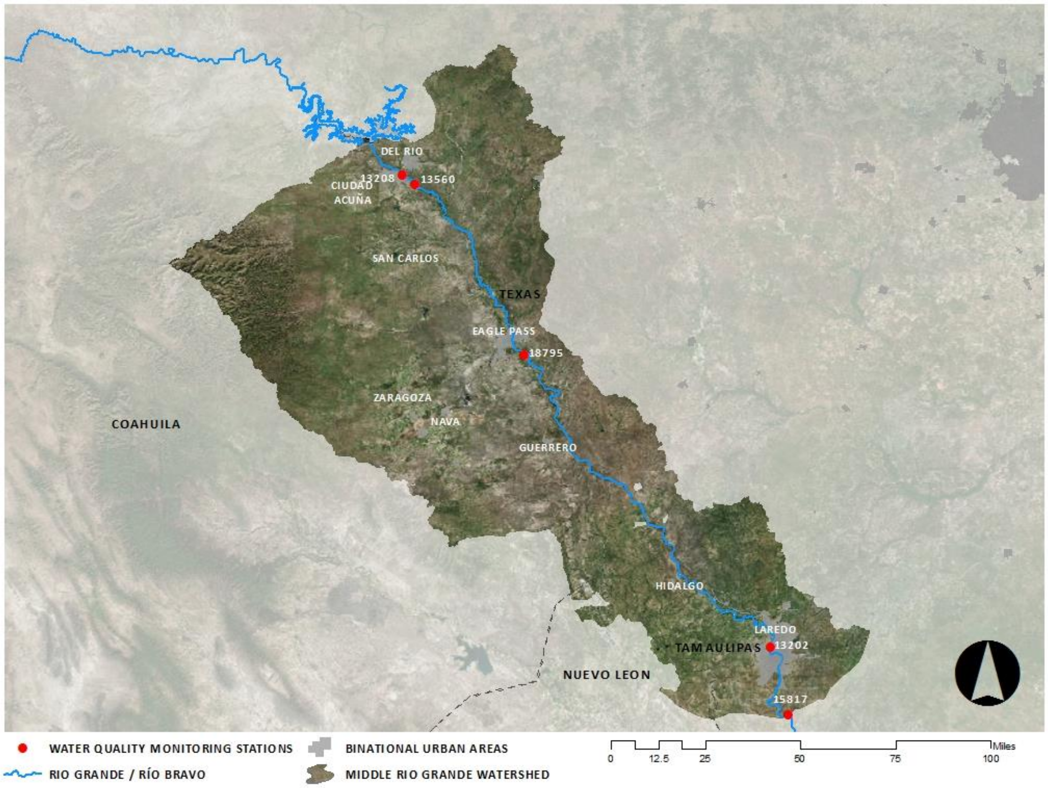
There are three classified segments in the Upper Rio Grande, which are given by both TCEQ’s long description and segment number: the Rio Grande above International Dam (segment 2314), the Rio Grande below Riverside Diversion Dam (segment 2307), and the Rio Grande above International Amistad Dam (segment 2306). Within these segments, there are five SWQMIS water quality monitoring stations used for this report: the Rio Grande immediately upstream of the confluence with Anthony Drain east of La Tuna Prison near the state line (station 13276), the Rio Grande at Courchesne Bridge, 1.7 miles upstream from American dam (station 13272), the Rio Grande 2.4 miles upstream from the Río Conchos confluence (station 13230), the Rio Grande below the Río Conchos confluence near Presidio (station 13229), and the Rio Grande at the mouth of Santa Elena Canyon (station 13228) (Figure 3).

Figure 3: SWQMIS Monitoring Stations Along the Upper Rio Grande

Location and Description of Water Quality Monitors in the Middle Rio Grande

There is one classified segment in the Middle Rio Grande and it is defined by the TCEQ as the Rio Grande below International Amistad Reservoir (segment 2304). There are five SWQMIS water quality monitoring stations in the Middle Rio Grande that are used for this report: the Rio Grande 12.8 miles below Amistad Dam, near gage, 340m upstream of US 277 bridge in Del Rio (station 13208), the Rio Grande 4.5 miles downstream of Del Rio at Moody Ranch (station 13560), the Rio Grande at Kickapoo Reservation, 1.92 kilometers south and 2.02 kilometers west of RR 1021 at Maverick County Highway 523, south of Eagle Pass (station 18795), the Rio Grande Laredo water treatment plant pump intake (station 13202), and the Rio Grande at the Webb/Zapata county line (station 15817) (Figure 4).

Figure 4: SWQMIS Monitoring Stations along the Middle Rio Grande



Location and Description of Water Quality Monitors in the Middle Rio Grande

There is one classified segment in the Lower Rio Grande and it is defined by the TCEQ as the Rio Grande below International Falcón Reservoir (segment 2302). There are four SWQMIS water quality monitoring stations in the Lower Rio Grande that are used for this report: the Rio Grande below Río Alamo near Fronton (station 13186), the Rio Grande at SH 886 near Los Ebanos (station 13184), the Rio Grande International Bridge at US 281 at Hidalgo (station 13181), and the Rio Grande El Jardín pump station, at the low water dam, 300 feet below intake (station 13177) (Figure 5).

Figure 5: SWQMIS Monitoring Stations along the Lower Rio Grande

Geographic Information System (GIS) Data Layers

The GIS data layers used in this report were acquired from numerous sources. From the USGS U.S.-Mexico Border Environmental Health Initiative (BEHI), U.S. and Mexico datasets are provided that integrate transboundary data on places, boundaries, and hydrography. Data on transboundary places include a dataset on major binational cities and binational urban area extents. The BEHI major binational cities layer contains only major cities and pairs in the U.S.-Mexico Border Region as defined by the BEHI study area and its information source is the USGS National Atlas and the Instituto Nacional de Estadística (INEGI). The BEHI urban area extents layer contains urban area data for the U.S. and Mexico corresponding to the BEHI study area, its data on the U.S. side were obtained from the U.S. Census in 2006, the Texas data were obtained from the Texas Natural Resources Information System (TNRIS) in 2006, and data on the Mexico side were obtained from INEGI in 2006 from their 2000 AGEB Census data. Data on binational boundaries include the BEHI binational study area, state boundaries for states along the U.S.-Mexico border, as well as U.S. counties and Mexican municipios in the study area. The BEHI binational study area includes the 250k 8-digit HUCs from the USGS Water Resources National Spatial Data Infrastructure (NSDI) node and 1:250K watershed boundaries from INEGI. The international boundary between the U.S. and Mexico is consistent with the USGS states polygon dataset, and was digitized using orthoimagery from the 2004 National Agriculture Imagery Program (NAIP) and digital vector files from INEGI.

This report uses data regarding transboundary hydrography including a binational watershed of the Rio Grande / Río Bravo as well as a geo-database of binational major streams and rivers. The binational watershed dataset is a synthesis of the USGS 8 digit HUC boundaries from the USGS Water Resources NSDI Node and INEGI's 1:250K scale watershed cuenca boundaries. Because the binational watershed extends beyond the study area implemented by NAFTA (100 kilometers on each side of the international border), it was necessary to use geo-processing tools in ArcGIS to clip the binational watershed to the study area (see Figure 1, p. 22). For the binational streams and rivers dataset, U.S. streams are from the USGS National Hydrography Dataset (NHD) at the 1:100,000 scale. Mexican streams are from INEGI at the 1:250,000 scale. The integration and routing of the Rio Grande basin using the 1:100,000 and 1:250,000 data was performed by the Center for Research in Water Resources (CRWR) at the University of Texas at Austin.

Wastewater Expenditure Data

In order to measure the effects of capital expenditures made towards improving in-stream bacteria concentration, this study collected water and wastewater expenditure data from myriad sources. Binational, federal, and state capital expenditures made on wastewater infrastructure projects were compiled for Rio Grande and had to meet two main qualifications: (1) the expenditure must have been completed between 1995 and 2012 and (2) the project must be spatially located within one of the 14 sub-watersheds, ensuring its effect would be geographically bound to the drainage area of the water body. In total, 57 binational, federal, or state-funded wastewater projects were identified that met these two qualifications, equaling a total of $US 1.197 billion. Out of the 57 total projects, 44 were certified and capitalized by the BECC and the NADB, making up a total of $US 983.89 million. These infrastructure projects combined NADB loans, BEIF grants, and Community Assistance Program (CAP) grants. To keep inventory, the state (U.S. or Mexico), county (U.S. or Mexico), community/entity, program, project category, project title, date of completion, quarter of completion, and total cost were recorded for each project so that these data could be used in linear regression. Additionally, the BECC website provided an estimate of the benefiting population for each project. Benefiting population may serve as both a proxy for population density as well as a benchmark for gauging a project's initial impact. Lastly, the BECC online database gives a geo-location for each of its projects, so it is possible to pinpoint which of the fourteen watersheds is being affected by each wastewater infrastructure project on a map.

In the State of Texas, the Texas Water Development Board (TWDB) provides funding for wastewater infrastructure projects in the border region as part of the Economically Distressed Areas Program (EDAP), the Clean Water State Revolving Fund (CWSRF), the Colonias

Wastewater Treatment Assistance Program (CWTAP), the Drinking Water State Revolving Fund (DWSRF), and the Water Development Fund (WDF). In this study, funding from the TWDB comprises 13 of the total 57 projects from 1995-2012 that affect water quality in the Rio Grande watershed. The TWDB's contribution to wastewater infrastructure is much less than those at the binational and federal levels, comprising approximately $US 213.21 million. Similarly to the BECC and NADB's estimations, data on benefiting population and project location were readily available through the TWDB online database.

For the purposes of this study, the baseline year for wastewater infrastructure projects is 1995. Any projects completed during or before 1994 are not included in the model. By using the date in which a wastewater project was completed, expenditures (in $US millions) were assigned to both the corresponding quarter and water quality monitoring station. For each water quality monitoring station, wastewater expenditures were collected and added accumulatively for each quarter from 1995 to 2012. Similarly, data concerning benefiting population for the appropriate quarter and water quality monitoring station were also added accumulatively for each quarter from 1995 to 2012.

Precipitation Data

This study uses daily rainfall totals obtained from the National Oceanic and Atmospheric Administration (NOAA) from Next-Generation Radar (NEXRAD) technology. NEXRAD is a high-resolution Doppler weather radar that can detect precipitation, atmospheric movement and wind. NEXRAD precipitation data are available on a daily basis but must be compiled using a Cygwin interface for data extraction. The data are extracted into 3,000 by 3,000 meter quadrants, given latitude and longitude coordinates specified in NEXRAD extraction process. For the purposes of this report, this allowed the researcher to pinpoint the quadrants surrounding the 14 water quality monitoring stations to get the most accurate precipitation measurements that could affect water quality at those locations (Figure 6). Daily precipitation data are then summed to a quarterly level in order to temporally synchronize with the water quality, expenditure, and benefiting population data.

Figure 6: NEXRAD Precipitation Quadrants in the Lower Rio Grande

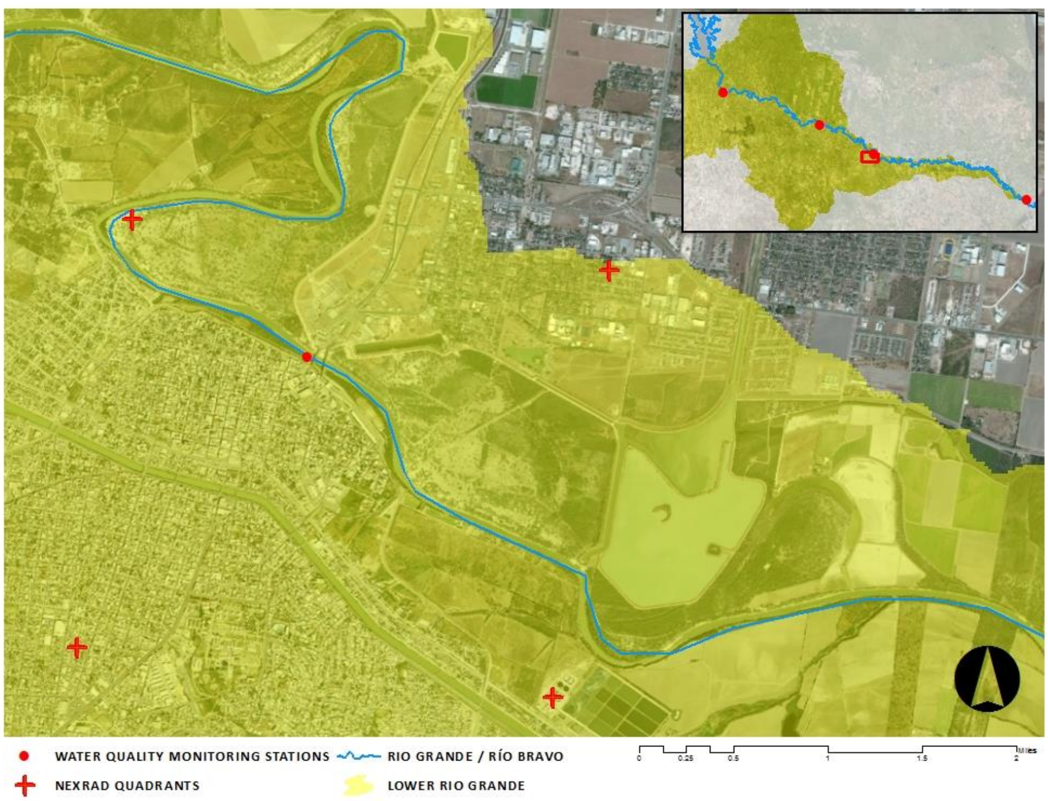


Table 5: Descriptive Statistics of Variables Used for Analysis

Variables[1]	Upper Rio Grande				Middle Rio Grande				Lower Rio Grande			
	Min	Max	Mean	N**	Min	Max	Mean	N**	Min	Max	Mean	N**
E. Coli Concentration (MPN cfu/100mL)	14	3,728	641	72	1	10,002	141	72	3	2,400	100	72
Quarterly Waste Water Expenditures (millions $US)	0.16	100.25	12.52	32	3.53	105.34	35.14	14	1.1	83.4	35.14	16
Quarterly Benefited Population	277	1,217,818	129,385	32	1,128	384,033	64,080	14	2,635	420,463	117,334	16
Quarterly Precipitation (mm)	0.00	304.80	120.71	72	0	137.16	23.95	72	0	655.83	127.28	72

[1]All values for the period 1995 - 2012
[*]Dependent variable
[**]Number of individual measurements
[^]Total Number of Projects

Table 5 gives an overview of the data used in this study, including information on water quality, wastewater expenditures, the benefiting population for each wastewater infrastructure project, and the quarterly precipitation for each of the three river sections. The Upper Rio Grande is, on average, the most polluted section of the river, even though the Middle Rio Grande exhibits the quarter with the maximum *E. coli* concentration in the study period. Between 1995 and 2012, average concentrations in both the Upper and Middle Rio Grande exceed the EPA 126 cfu/100mL criterion for freshwater contact recreation, while the Lower Rio Grande's average concentration is just below this criterion. The number of wastewater infrastructure projects has been most concentrated in the Upper Rio Grande. The Upper Rio Grande has nearly three times as many projects as the Middle Rio Grande and Lower Rio Grande sections. However, the average project cost in both the Middle and Lower Rio Grande is nearly three times the average project cost in the Upper Rio Grande. On balance, expenditures are lower on average in the Upper Rio Grande with the other two sections approximately equal. Project cost differences aside, the benefiting population in the Upper Rio Grande, on average, exceeds both the Middle and Lower Rio Grande sections. The low average in benefiting population in the Middle Rio Grande is consistent with a lower population density in this area, compared to the Upper and Lower Rio Grande sections. Unsurprisingly, all three sections show low precipitation levels that are indicative of a semi-arid to arid climate. The Upper and Lower Rio Grande experience similar levels of rainfall, on average, though the Lowe Rio Grande often experiences more pronounced storm events from hurricanes and tropical storms. The Middle Rio Grande is the driest of the three sections.

CHAPTER 4: ANALYTICAL METHODS

To estimate the impact of capital investment in wastewater infrastructure on in-stream bacterial concentration in each of the Rio Grande sub-watersheds, this study utilizes a combination of analytical methods. To demonstrate the iterative process of applying these methods, this chapter is divided into five sub-sections: addressing missing data in this study, using multiple imputation to estimate data missingness, using a panel data linear regression approach for program evaluation, deciding fixed versus random effects in the linear regression model, and model specification based on variables included in this study.

Missing Data

Due to a number of factors, missing data are common in multivariate water quality sampling, especially in longitudinal studies encompassing panel data from different intercepts (water quality monitoring stations) over time (Hirsch, 1982). State, federal, or multilateral funding mechanisms may not be sufficient for quarterly or even yearly water quality sampling for a given monitoring station, and sampling efforts may not be carried out due to the burden of the lack of funding, the difficulty of meeting quality assurance planning, or other bureaucratic constraints. The simplest way to treat missing data is to delete them. This method, known as list wise deletion, is the default option in almost all statistical software packages. However, doing so produces biased inferences and reduces statistical power in multivariate modeling (Allison, 2002). Another approach used to treat missing data is single imputation, either using a simple mean, an adjusted mean from a regression model, or a “hot deck” procedure. This approach offers a complete dataset and may support unbiased inferences. However, single imputation does not provide valid standard error calculations and confidence intervals that account for the uncertainty caused by the imputed missing data. Under certain assumptions, multiple imputation can be employed to provide valid references from a complex and incomplete dataset that account for the uncertainty involved in imputing missing values.

In order to generate meaningful results from longitudinal regression, it was imperative for this study to organize the dependent variable (in-stream bacteria concentration) so that data would be present for each of the 72 quarters in the study from 1995 to 2012. However, complications arose from the widespread missingness of *E. coli* on a quarterly basis (See Table 6).

Table 6: *E. coli* Data Patterns of Missingness in the Rio Grande / Río Bravo, 1995-2012

River Section	**Station ID**	**Watershed Number**	**Quarters of Missing Data^**	**Completeness of Data***
Upper Rio Grande	13276	1	22	69%
Upper Rio Grande	13272	2	27	63%
Upper Rio Grande	13230	3	28	61%
Upper Rio Grande	13229	4	29	60%
Upper Rio Grande	13228	5	36	58%
Middle Rio Grande	13208	6	36	58%
Middle Rio Grande	13560	7	29	60%
Middle Rio Grande	18795	8	27	63%
Middle Rio Grande	13202	9	0	100%
Middle Rio Grande	15817	10	29	60%
Lower Rio Grande	13186	11	46	36%
Lower Rio Grande	13184	12	58	19%
Lower Rio Grande	13181	13	49	32%
Lower Rio Grande	13177	14	49	32%

^Out of the 72 quarters from 1995 - 2012
*A complete dataset would be at the 100% level

For all 14 water quality monitoring stations used in this report, data for the dependent variable, *E. coli*, was missing an average 55% of the time. For the entire study area, water quality monitoring stations were missing *E. coli* data for an average of 33 quarters (out of 72 quarters). Some river sections exhibited higher degrees of data missingness than others, however. On average, the Lower Rio Grande's four water quality monitoring stations showed 30%

completeness of data (70% missingness), worse than both the Middle Rio Grande (66% completeness; 34% missingness) and the Upper Rio Grande (64% completeness; 36% missingness). Out of all 14 water quality monitoring stations, station 13202 in the Middle Rio Grande was the only station that had complete data. The water quality monitoring station with the least amount of data present was station 13184, which reported *E. coli* data only 19% of the time during the study period. Given the prevalence of missing data, it was necessary to explore alternative methods for estimating these missing values.

Multiple Imputation

Multiple imputation, proposed by Rubin (1987), has emerged as a flexible alternative to likelihood methods for a wide variety of missing-data problems. Multiple imputation (MI) is a general statistical method for the analysis of incomplete data sets. The three basic steps of a statistical analysis using multiple imputation begin with specifying and generating plausible synthetic data values, called imputations, for the missing values in the data. This step results in a number of complete data sets (m) in which the missing data are replaced by random draws of $m > 1$ simulated values, substituting the j^{th} element of each list for the corresponding missing value, where $j = 1, \dots, m$, producing m plausible alternative versions of the complete data. The second step consists of analyzing each imputed data set by a statistical method that will estimate the quantities of interest, resulting in m analyses. The third step pools the m estimates into one estimate, thereby combining the variation within and across the m imputed data sets. For a theoretical definition of multiple imputation and demonstration of the multiple imputation procedure, please see the Appendix.

Panel Data

Panel data refers to multi-dimensional data that contain observations of multiple phenomena obtained over multiple time periods of time for the same cross-section or intercept (Kennedy, 2008). One of the main attributes of panel data regression is that it can enable correction of the problem of heterogeneity between the intercepts of a cross-sectional or longitudinal model. In any intercept, myriad unmeasured explanatory variables may affect the behavior of the cross-section. In this report, the intercepts are confined to each of the 14 water quality monitoring stations to be analyzed. Heterogeneity, in this case, means that each of the water quality monitoring stations is different from the others in fundamental unmeasured ways over time. Omitting these variables causes bias in estimation, but panel data models allow each of the water quality monitoring stations to have different intercepts, thus solving the heterogeneity problem in a linear regression. Panel data regression is also useful in creating more variability, studying trends over time, and allowing better analysis of dynamic adjustment (Kennedy, 2008). A general panel data regression model is written as:

$$y_{it} = \alpha + \beta' X_{it} + u_i + \varepsilon_{it} \tag{1}$$

where y_{it} is the dependent variable at monitor i in quarter t, X_{it} are the independent variables, β are the coefficients that describe the size of the effect the independent variables have on the dependent variable, and α is a constant. The error structure comprises u_i, the monitor effect (which may either be fixed or random), and ε_{it}, the standard Gauss-Markov error term.

Fixed Effects versus Random Effects

A fixed effects (FE) panel data model is used when a researcher is only interested in analyzing the impact of variables that vary over time. According to Allison (2009), fixed effects explore the relationship between predictor and outcome variables within an entity and each entity has its own individual characteristics that may or may not influence the predictor variables. In this study, a FE model would assume that a water quality monitoring station would have individual characteristics (e.g. site-specific environmental infrastructure projects, site-specific rainfall events, etc.) that influence the ability to predict the outcome variable of in-stream bacterial concentration. A FE model removes the effect of time-invariant characteristics from the estimates so it is possible to assess the net effect of the predictors on the outcome of the dependent variable. Another important assumption of the FE model is that characteristics that are time-invariant unique to the site effect u_i should not be correlated with other individual characteristics. If the error terms in an intercept are correlated, then FE is not suitable; inferences may be biased and the research may consider re-modeling the relationship. This is the main rationale for the Hausman test (Allison, 2009).

A random effects (RE) model differs from an FE model in that, unlike an FE model, the variation across intercepts is assumed to be random and uncorrelated with the predictor or independent variables included in the model:

> "…the crucial distinction between fixed effects and random effects is whether the observed individual effect embodies elements that are correlated with the regressors in the model, not whether these effects are stochastic or not." (Greene, 2008 p. 183)

If there a reason to suspect that differences across intercepts may hold influence on the dependent variable, then RE should be used as the model specification. An RE model assumes that the error term is not correlated with the predictors and allows for time-invariant variables to play a role as explanatory variables.

To decide between fixed or random effects, researchers often rely on the Hausman (1978) specification test (Greene 2008, pp. 208-209). The Hausman test is designed to detect violation of the RE modeling assumption that the explanatory variables are orthogonal to the unit effects (Hausman, 1978). If there is no correlation between the independent variable(s) and the unit effects, then the estimates of β in the FE model $(\hat{\beta}_{FE})$ should be similar to the estimates of β in the random effects model $(\hat{\beta}_{RE})$. The Hausman test statistic H is a measure of the difference between the two estimates:

$$H = (\hat{\beta}_{RE} - \hat{\beta}_{FE})' \left[\mathrm{Var}(\hat{\beta}_{FE}) - \mathrm{Var}(\hat{\beta}_{RE}) \right]^{-1} (\hat{\beta}_{RE} - \hat{\beta}_{FE}) \tag{2}$$

Under the null hypothesis of orthogonality, H is distributed chi-square with degrees of freedom equal to the number of regressors in the model. A finding that $p < 0.05$ is taken as evidence that the two models are *different enough* to reject the null hypothesis, and thus to reject the random effects model in favor of the fixed effects model (Hausman, 1978).

Model Specification

The panel data models estimated here take the form:

$$Y_{it} = \beta_1 X_{1,it} + \beta_2 X_{2,it} + \beta_3 X_{3,it} + \beta_4 X_{4,it} + u_i + \varepsilon_{it} \tag{3}$$

Where Y_{it} is *E. coli* concentration (MPN/100mL) at monitor i in quarter t, $X_{1,it}$ represents quarterly wastewater expenditures (in \$US millions), $X_{2,it}$ is the benefiting population of each wastewater infrastructure project, $X_{3,it}$ is the ratio of expenditures divided by benefiting population, capturing the level of investment per capita, $X_{4,it}$ is quarterly rainfall in millimeters, u_i $(i = 1...n)$ is the time-invariant fixed effect for each monitor, and ε_{it} is the error term. The terms $X_{1,it}$ and $X_{2,it}$ each represent accumulative totals per monitor per quarter. Unless there is expenditure made in the first quarter of 1995, each monitor begins at zero $US million. Over time, each monitor quarter displays the total $US millions spent in that quarter in addition to any expenditures made in the previous monitor quarter(s) so that, by December 2012, $X_{1,it}$ and $X_{2,it}$ show all expenditures made between the first quarter of 1995, and the last quarter of 2012.

CHAPTER 5: RESULTS AND RECOMMENDATIONS

Standard Hausman Test

The result of the standard Hausman test reported in Table 7 indicates that the orthogonality hypothesis of the unobservable individual-specific effects and the regressors is rejected. In each of the three river sections, the *p* values in the $p > chi^2$ test denote significance at the 1% level (Table 7). Thus, for each of the river sections, the two models are different enough to reject the null hypothesis, rejecting the RE model in favor of FE.

Table 7: Standard Hausman Test Results for Fixed and Random Effects

River Section	Variable	Fixed Effects	Random Effects
Upper Rio Grande	*Expenditures*	-0.103** (1.017)	0.610 (0.867)
	Precipitation	-0.017** (0.571)	0.770 (0.591)
Middle Rio Grande	*Expenditures*	-1.388** (12.58)	-29.09 (10.01)
	Precipitation	0.237** (0.846)	0.194 (0.877)
Lower Rio Grande	*Expenditures*	1.698** (1.050)	-2.934 (0.911)
	Precipitation	0.165** (0.154)	0.132 (0.171)

**Denotes significance at the 1% level.

First Stage Fixed Effects Panel Data Regression Results

In order to test if the multiple imputation process made an observable difference, we run an FE model containing missing data (noted as Missing FE in Tables 8, 9, and 10) along with a second FE model containing multiply-imputed values for *E. coli* for each of the monitoring stations (noted as Model 2 FE in Tables 8, 9, and 10). Additionally, we run a third FE model that uses multiply-imputed values and also includes an interaction variable to sufficiently control for any underlying heterogeneity in pollution concentration trends over time (noted as Model 3 FE in Tables 8, 9, and 10). The interaction variable is generated by multiplying the monitor effects times FEs created to represent each year.

Upper Rio Grande

The model estimates for in-stream *E. coli* concentration are generally consistent, showing mostly statistically insignificant relationships for each of the explanatory variables (Table 8). In the Missing FE model (column 1), the *Expenditures* coefficient is positive, significant at the 10% level, and is much higher in magnitude than the Model 2 FE and Model 3 FE *Expenditures* coefficients. *Benefitting Population* is also significant at the 10% level in the Missing FE model, its coefficient is negative, and appears to have a very small effect on *E. coli* concentration. The R-squared value for the Missing FE model is about one-quarter that of Model 2 FE, suggesting that the model containing the multiply-imputed data possesses more explanatory power. Covariates are slightly more robust and conform better to expectations under Model 2 FE, with a noticeable change in magnitude of the *Expenditures* coefficient from positive to negative. The difference in *Expenditures / Benefiting Population* coefficients between Model 2 FE and Missing FE appears small but, unlike the Missing FE model, the coefficient in Model 2 FE is significant at the 10% level. The use of the Monitor-Year FE interaction term in Model 3 FE changes the coefficients from negative to positive for *Expenditures* and from positive to negative for *Precipitation.* The most noticeable difference, however, from Model 2 FE to Model 3 FE is that the *Expenditures / Benefiting Population* variable becomes significant at the 1% level in Model 3 FE. Its positive coefficient suggests that increased spending on wastewater projects in the region increases bacteria concentration by a small margin. One explanation of why this result does not conform to expectation may be that Model 3 FE still does not sufficiently control for underlying heterogeneity.

Table 8: Upper Rio Grande First Stage Fixed Effects Panel Data Results

E. coli	(1) Missing FE	(2) Model 2 FE	(3) Model 3 FE
Expenditures	9.38529^{+} (5.2718)	-0.10931 (0.9354)	0.49544 (1.0078)
Benefiting Population	-0.00219^{+} (0.0011)	-0.00003 (0.0001)	-0.00006 (0.0001)
Expenditures / Benefiting Population	0.00643 (0.0107)	0.01117^{+} (0.0063)	0.02237** (0.0081)
Precipitation	0.23704 (0.7757)	0.04077 (0.4633)	-0.09736 (0.4642)
Monitor FEs	Yes	Yes	Yes
Monitor -Year FEs	No	No	Yes
R^2	0.0861	0.3203	0.4022
Observations (N)	218	360	360
Monitors (*i*)	5	5	5
Quarters (*t*)	72	72	72

Notes: **Denotes significance at the 1%-level. *Denotes significance at the 5%-level. +Denotes significance at the 10%-level. The dependent variable is *E. coli* concentration (MPN / 100mL).

Middle Rio Grande

In the Middle Rio Grande section, all model estimates are statistically insignificant in affecting the dependent variable, *E. coli*. Although the magnitude of the *Expenditures* coefficient is reduced in Model 2 FE, there appears to be little remaining difference between the Missing FE model and Model 2 FE coefficient estimates. Despite the inclusion of multiply-imputed data, the R-squared value in Model 2 FE decreases by almost two-thirds. The only observable result from including the Monitor-Year FE interaction term in Model 3 FE is that the coefficient for the *Expenditures* variable becomes negative. Otherwise, coefficients stay relatively similar between the two models but the R-squared value more than doubles from Model 2 FE to Model 3FE.

Table 9: Middle Rio Grande First Stage Fixed Effects Panel Data Results

E. coli	(1) Missing FE	(2) Model 2 FE	(3) Model 3 FE
Expenditures	4.10607	0.81849	-0.68429
	(5.7413)	(1.2647)	(1.6686)
Benefiting Population	-0.00040	-0.00020	-0.00070
	(0.0015)	(0.0005)	(0.0006)
Expenditures / Benefiting Population	-0.01333	-0.00586	-0.01820
	(0.0784)	(0.0340)	(0.0353)
Precipitation	-2.85889	-1.95095	-1.97842
	(2.6521)	(1.9043)	(1.9036)
Monitor FEs	Yes	Yes	Yes
Monitor -Year FEs	No	No	Yes
R^2	0.1854	0.0658	0.1625
Observations (N)	238	360	360
Monitors (*i*)	5	5	5
Quarters (*t*)	72	72	72

Notes: **Denotes significance at the 1%-level. *Denotes significance at the 5%-level. +Denotes significance at the 10%-level.
The dependent variable is *E. coli* concentration (MPN / 100mL)

Lower Rio Grande

All three model estimates show statistically significant results for *Benefiting Population*'s effect on *E. coli* concentration in the Lower Rio Grande. The Missing FE model shows *Expenditures* is significant at the 1% level but its coefficient is positive and large when compared to the estimates of the other two models. The coefficient for *Expenditures* becomes much smaller when multiply-imputed data are used in Model 2 FE and the *Expenditures* coefficient is negative in Model 3 FE, with the inclusion of the Monitor – Year FE interaction term. In comparing R-squared values, Model 2 FE has more than twice the explanatory power of the Missing FE model. The number of observations between these two models changes drastically, given the significant missingness of water quality data in the Lower Rio Grande, with the Missing FE model using 85 observations compared to Model 2 FE's 288 observations. Aside from the change in the *Expenditure* coefficient between Model 2 FE and Model 3 FE, Model 3 FE also shows a statistically significant relationship at the 1% level between the *Benefiting Population* and *E. coli* concentration. For every 10,000 people per quarter that are benefiting from a wastewater project in the Lower Rio Grande, *E. coli* can be expected to decrease by about 8 colony forming units per quarter year. The change in the statistical significance of *Benefiting Population* at the 5% level in Model 2 FE to statistical significance at the 1% level in Model 3 FE may be related to the fact the R-squared value increases almost 50% from Model 2 FE to Model 3 FE.

Table 10: Lower Rio Grande First Stage Fixed Effects Panel Data Results

E. coli	(1) Missing FE	(2) Model 2 FE	(3) Model 3 FE
Expenditures	19.0473**	1.69829	-0.28724
	(5.8052)	(1.0496)	(1.2295)
Benefiting Population	-0.00202*	-0.00048*	-0.00075**
	(0.0010)	(0.0002)	(0.0003)
Expenditures / Benefiting Population	-0.01163	-0.02520	-0.01330
	(0.2841)	(0.0542)	(0.0580)
Precipitation	-0.31229	0.16476	0.19943
	(0.3730)	(0.1542)	(0.1523)
Monitor FEs	Yes	Yes	Yes
Monitor -Year FEs	No	No	Yes
R^2	0.1263	0.2810	0.4044
Observations (N)	85	288	288
Monitors (*i*)	4	4	4
Quarters (*t*)	72	72	72

Notes: **Denotes significance at the 1%-level. *Denotes significance at the 5%-level. +Denotes significance at the 10%-level. The dependent variable is *E. coli* concentration (MPN / 100mL)

Comparison of Models Across Regions

Overall, estimates are mixed from each region in showing explanatory variables' effect on *E. coli* concentration. One trend consistent throughout each section's *Expenditures* coefficients is the progression from a positive to a negative coefficient, as the model specification and data quality are improved, with exception of the outlier present in the Upper

Rio Grande Model 3 FE. Conversely, *Precipitation* coefficients behave in the opposite manner by mostly starting off as negative coefficients in the Missing FE models and gradually becoming positive value. Both of these changes in coefficients move toward hypothesized outcomes and suggest that the use of multiply-imputed data and Monitor-Year FE terms improves the explanatory power of statistical inference. Lastly, it is clear from all three sections that precipitation does not have a statistically significant relationship to in-stream bacteria concentration, once we control for average concentrations at a monitor (u_i) and in a monitor year.

Second Stage Fixed Effects Panel Data Regression Results

In order to further explore the relationship between each of the independent variables' effect on *E. coli* concentration in the Rio Grande, it was necessary to run a second stage of regressions using panel data. This stage uses Model 3 FE but runs each independent variable (*Expenditures, Benefiting Population, and Expenditures / Benefiting Population*) in a separate regression. These variables are all collinear, thus it is, perhaps, not surprising that few of their coefficients are significant when all three are included in the models. Results are provided by Table 11, 12, and 13.

Upper Rio Grande

Model estimates show significant results at the 1% level only when isolating *Expenditures / Benefiting Population* (Table 12). Although negligible, its coefficient is positive, suggesting that running this variable in a separate regression may not fully control for underlying heterogeneity. The R-squared value is nearly equal for all three models. Both the *Expenditures* and *Benefiting Population* variables' coefficients are negative and insignificant.

Table 11: Upper Rio Grande Second Stage Fixed Effects Panel Data Results

E. coli	(1) Model 3 *Expenditures*	(2) Model 3 *Benefiting Population*	(3) Model 3 *Expenditures / Benefiting Population*
Expenditures	-0.00081 (0.4638)		
Benefiting Population		-0.00002 (0.0001)	
Expenditures / Benefiting Population			0.02239** (0.0079)
R^2	0.4067	0.4088	0.4027
Observations (N)	360	360	360
Monitors (*i*)	5	5	5
Quarters (*t*)	72	72	72

Notes: **Denotes significance at the 1%-level; * denotes significance at the 5%-level; +denotes significance at the 10%-level. The dependent variable is *E. coli* concentration (MPN / 100mL). All models include monitor and monitor-year fixed effects.

Middle Rio Grande

As in Table 9, model estimates for the Middle Rio Grande show no statistically significant results (Table 12). Coefficients for *Expenditures* and *Benefiting Population* are both negative. All three models have to low R-squared values and do not exhibit robust explanatory power.

Table 12: Middle Rio Grande Second Stage Fixed Effects Panel Data Results

E. coli	(1) Model 3 *Expenditures*	(2) Model 3 *Benefiting Population*	(3) Model 3 *Expenditures / Benefiting Population*
Expenditures	-0.94267 (1.6357)		
Benefiting Population		-0.00050 (0.0005)	
Expenditures / Benefiting Population			0.00423 (0.0302)
R^2	0.1618	0.1626	0.1627
Observations (N)	360	360	360
Monitors (*i*)	5	5	5
Quarters (*t*)	72	72	72

Notes: **Denotes significance at the 1%-level; * denotes significance at the 5%-level; +denotes significance at the 10%-level. The dependent variable is *E. coli* concentration (MPN / 100mL). All models include monitor and monitor-year fixed effects.

Lower Rio Grande

Model estimates show that *Expenditures* and *Benefiting Population* are both statistically significant in reducing bacteria concentration in the Lower Rio Grande (Table 13). *Expenditures* is significant at the 5% level, for every $US 1 million spent on wastewater infrastructure, *E. coli* levels will decrease by about 2 colony-forming units in the Lower Rio Grande watershed. *Benefiting Population* is significant at the 1% level and for every 10,000 people benefiting from a wastewater project, *E. coli* concentration can be expected to decrease by 7.6 colony-forming units in the region. Although *Expenditures / Benefiting Population* is not statistically significant, its coefficient is negative. All three variables show fairly high R-squared values when compared to R-squared values from the Lower Rio Grande in the first stage fixed effects panel data regression.

Table 13: Lower Rio Grande Second Stage Fixed Effects Panel Data Results

E. coli	(1) Model 3 *Expenditures*	(2) Model 3 *Benefiting Population*	(3) Model 3 *Expenditures / Benefiting Population*
Expenditures	-2.32585* (0.9633)		
Benefiting Population		-0.00076** (0.0002)	
Expenditures / Benefiting Population			-0.01289 (0.0570)
R^2	0.4047	0.4043	0.4098
Observations (N)	288	288	288
Monitors (*i*)	4	4	4
Quarters (*t*)	72	72	72

Notes: **Denotes significance at the 1%-level; * denotes significance at the 5%-level; +denotes significance at the 10%-level. The dependent variable is *E. coli* concentration (MPN / 100mL). All models include monitor and monitor-year fixed effects.

Overall, these models have been most effective by revealing the statistical significance of key explanatory variables within the Lower Rio Grande watershed.

Pooled Fixed Effects Panel Data Regression

Table 14 offers model estimates that pools together all observations for the Upper, Middle, and Lower Rio Grande river sections. In column 1, Model 3 FE includes all of the key explanatory variables (*Expenditures*, *Benefiting Population*, *Expenditures / Benefiting Population, and Precipitation*) in one FE regression. In column 2, Model 3 *Expenditures* FE isolates *Expenditures* as an explanatory variable as an attempt to minimize collinearity present in Model 3 FE. Model 3 *Benefiting Population* FE in column 3 isolates *Benefiting Populations*

Table 14: Pooled First Stage and Second Stage Fixed Effects Panel Data Results

E. coli	(1) Model 3 FE	(2) Model 3 *Expenditures* FE	(3) Model 3 *Benefiting Population* FE	(4) Model 3 *Expenditures / Benefiting Population* FE
Expenditures	-0.76758 (0.7482)	-0.44767 (0.4928)		
Benefiting Population	-0.00007 (0.0001)		-0.00009 (0.0001)	
Expenditures / Benefiting Population	0.00744 (0.0091)			0.00903 (0.0090)
Precipitation	0.07006 (0.2390)			
R^2	0.0936	0.0931	0.0931	0.0937
Observations (N)	1008	1008	1008	1008
Monitors (*i*)	14	14	14	14
Quarters (*t*)	72	72	72	72

Notes: **Denotes significance at the 1%-level; * denotes significance at the 5%-level; +denotes significance at the 10%-level. The dependent variable is *E. coli* concentration (MPN / 100mL). All models include monitor and monitor-year fixed effects.

as an explanatory variable while Model 3 *Expenditures / Benefiting Population* FE in column 4 isolates *Expenditures / Benefiting Population* as an explanatory variable. When all three river section's 1008 observations are pooled together, none of the models provide statistically significant results. In line with the hypothesized outcomes of how expenditures and benefiting population affect bacterial concentration in a water body, the coefficients for *Expenditures* and *Benefiting Population* are negative in both the Model 3 FE and isolated results. Coefficients for *Expenditures / Benefiting Population* are positive in both the Model 3 FE and isolated results, suggesting that an increase in expenditures per person in the Rio Grande watershed has a small but positive affect on in-stream bacteria concentration. Because none of the results are statistically significant and because the R-squared values in these four models have only one fourth of the explanatory power than from sectional models (Tables 8-10 and 11-13 respectively), it is not possible to draw strong inferences from these results. Even though the *Precipitation* coefficient conforms to the hypothesized outcome of having a positive value, it is also statistically insignificant in Model 3 FE.

Recommendations

While the econometric models do not perform as well as hoped, important takeaways can be gleaned from the above results. Both public expenditures on wastewater infrastructure and the size of the benefiting population reduce *E. coli* concentrations significantly in the Lower Rio Grande watershed. While the model tests provide these hints, they do not conclusively demonstrate a causal relationship between investment in wastewater infrastructure and bacteria concentration in the Rio Grande. However, we can say with confidence that there is a significant relationship between the wastewater expenditures and *E. coli* levels in the Lower Rio Grande. This should inform future decision-making regarding continued investment in the region. In addition to prioritizing wastewater infrastructure projects in the Lower Rio Grande region, it is also noteworthy to emphasize the effectiveness of larger wastewater projects in this section. Investments on projects that impact a larger number of people are shown to significantly impact the reduction of bacterial water pollution in the Lower Rio Grande. Further research is required to determine why these effects are not detectable in the other sections of the river.

This study has also highlighted the importance of collecting robust data. Models estimated on the original sample, with a high percentage of missing observation, provided weak explanatory power attributable to the lack of quarterly data collected on bacteria levels for monitoring stations along the Rio Grande / Río Bravo. With the inclusion of multiply-imputed values for missing data, the models offer more accurate and powerful statistical inferences. The use of multiple imputation has proven to be effective for this project, but a more extensive collection of quality-assured data would allow for more reliable analysis in future studies.

Lastly, further research is necessary to more extensively identify and explain the myriad factors that influence bacteria concentration in the Rio Grande. The models in this study primarily focused on the effect of wastewater expenditures on *E. coli* levels in the Rio Grande. Data describing land use, permitted wastewater discharge, non-point source pollution as the result of animal grazing, and population growth were not included in this report. These influences are captured by the flexible controls in the models – the monitor and monitor-year FEs. However, given this "black box" approach to controlling for potentially confounding sources of water quality variation – which reduces omitted variables bias in identifying the impact of wastewater treatment expenditures – we cannot determine from the models whether these additional pollution sources are more, less, or equal contributors to bacterial concentrations in the Rio Grande. Further research on this topic is warranted in order to determine the most effective policy approaches to improving the river's water quality.

APPENDIX

Multiple Imputation Method Notation

Let Y_j be one of k incomplete random variables (j = 1, … , k) and let $Y = (Y_1, \ldots, Y_,)$. The observed and missing parts of Y_j are denoted by Y^{obs} *and* Y_j^{mis}, respectively, so $Y^{obs} = (Y^{obs}, \ldots, Y_{,k}^{obs})$ *and* $Y^{mis} = (Y_1^{mis}, \ldots, Y_k^{mis})$ stand for the observed and missing data in Y. *Let* $Y_{-j} = (Y_1, .., Y_{j-1}, Y_{j+1}, \ldots, Y_k)$ denote the collection of the $k - 1$ variables in Y except Y_j. Let Rj be the response indicator of Yj, with $R_j = 1$ if Y_j is observed and $R_j = 0$ if Y_j is missing. Let $R = (R_1, \ldots, R_k)$ and $R_{-j} = (R_1, \ldots, R_{j-1}, R_{j+1}, \ldots, R_k)$. Let $X = (X_1, \ldots, X_l)$ be a set of l complete covariates on the same subjects. In order to avoid complexities, it is assumed that the observations in Y, X, and R correspond to a simpler random sample from the population of interest.

The simplest method for combining the results of m analyses is Rubin's (1987) method for a scalar (one-dimensional) parameter. In this study, Q represents a bacteria concentration quantity (e.g. a regression coefficient) to be estimated. Let $\hat{Q}$ and $\sqrt{U}$ denote the estimate of Q and the standard error that one would use if no data were missing. The method assumes that the sample is large enough so that $\sqrt{U}(\hat{Q} - Q)$ has approximately a standard normal distribution, so that $\hat{Q} \pm 1.96\sqrt{U}$ has about 95% coverage. We cannot compute $\hat{Q}$ and U ; rather, we have m different versions of them $\lfloor \hat{Q}^{(j)}, U^{(j)} \rfloor, j = 1, \ldots, m$. Rubin's (1987) overall estimate is simply the average of the m estimates,

$$\overline{Q} = m^{-1} \sum_{j=1}^{m} \hat{Q}^{(j)}$$

The uncertainty of $\overline{Q}$ has two parts: the average within-imputation variance,

$$\overline{U} = m^{-1} \sum_{j=1}^{m} U^{(j)}$$

And the between-imputations variance,

$$B = (m-1)^{-1} \sum_{j=1}^{m} \left[\hat{Q}^{(j)} - \overline{Q}\right]^2$$

The total variance is a modified sum of the two components,

$$T = \overline{U} + (1 + m^{-1})B$$

And the square root of *T* is the overall standard error. For confidence intervals and tests, Rubin (1987) recommended the use of a *t* approximation, $T^{-1/2}(\overline{Q} - Q) \sim t_v$, where the degrees of freedom are given by

$$v = (m-1)\left[1 + \frac{\overline{U}}{(1 + m^{-1})B}\right]^2$$

The degree of freedom may vary from *m* – 1 to ∞ depending on the rate of missing information. When the degrees of freedom are large, the *t* distribution is essentially normal, the total variance is well estimated, and there is little to be gained by increasing *m*. The estimated rate of missing information for Q is approximately $\tau/(\tau+1)$, where $\tau = (1 + m^{-1})B/U$ is the relative increase in variance due to nonresponse. Additional methods for combining multidimensional parameter estimates, likelihood-ratio test statistics, and probability values from hypothesis tests are reviewed by Schafer (1997, chapter 4).

Multiple Imputation Procedure with Fully Conditional Specification

The validity of MI relies upon how the imputations are created and how that procedure relates to the model used to subsequently analyze the data. Creating MIs has greatly been helped by the advancement of computer software packages in specifying special algorithms (Schafer, 1997). Using SAS ® 9.3 Multiple Imputation procedure command (PROC MI), one is able to implement an algorithm given by Schafer (1997) for the parameters (means and co-variances) of the joint distribution of observed and missing variables in an iterative fashion, starting with the observed data and plausible values for the missing values. Fully Conditional Specification (FCS) allows imputation of data on a variable-by-variable basis by specifying an imputation model per variable. FCS is an attempt to define $P(Y, X, R \mid \theta)$ by specifying a conditional density $P(Y_j \mid X, Y_{-j}, R, \theta_j)$ for each Y_j. This density is used to impute Y_j^{mis} given X, Y_{-j} and R. Starting from simple guessed values, imputation under FCS is done by iterating over all conditionally specified imputation models. One iteration consists of one cycle through all Y_j. Analysis with multiple imputation is generally carried out in three steps. The first command for MI with FCS for one station is as follows:

Figure 7: SAS ® 9.3 Code for FCS Multiple Imputation

```
proc mi data=stationone out=miout1 nimpute=72; var spconduct chlor sulf
restotfilt nitamontot nitkjel secchidisc rstotnon nitrinitra fecalcol
bod5 ecoli; FCS regpmm(chlor sulf restotfilt nitamontot nitkjel
secchidisc rstotnon nitrinatra fecalcol bod5 ecoli); run;
```

With the imputation step, missing data are filled in using *m* different sets of values which produces *m* imputed datasets. To finish this step, one must examine the efficiency indicators in the output from PROC MI to determine whether enough imputations have been created. A key indicator of the efficiency of the imputation is the degrees of freedom (DF) for posterior parameter estimates. The DF is a direct function of efficiency. Allison (2001) has shown that variables with smaller DF appear to be those that have other problematic features including unsatisfactory internal consistency, many missing values, low incidence, and may be skewed.

The next step of the multiple imputation procedure is analysis, which is generated by using SAS® code shown in Figure 8:

Figure 8: SAS ® 9.3 Code for Multiple Imputation Analysis

```
proc mixed data=miout3; model ecoli= spconduct chlor sulf restotfilt
nitamontot nitkjel secchidisc rstotnon nitrinitra fecalcol bod5/solution
covb; by _Imputation_; ods output SolutionF=mixparms CovB=mixcovb; run;
```

Here, each of the m imputed datasets is analyzed separately using any method that would have been chosen had the data been complete.

Lastly, the analyzed dataset is pooled by using SAS ® code shown in Figure 9:

Figure 9: SAS ® 9.3 Code for Multiple Imputation Pooling

```
proc mianalyze parms(classvar=full)=mixparms; class ecoli; modeleffects
spconduct chlor sulf restotfilt nitamontot nitkjel secchidisc rstotnon
nitrinitra fecalcol bod5; run;
```

After the imputed data have been pooled, the overall estimate, as Rubin (1987) states, is simply the average of the *m* estimates. We may then take the imputed averages of *m* estimates and use them as our dependent variable in our longitudinal panel data regression.

Bibliography

Allison, Paul D. (2002) Missing Data: Quantitative Applications in the Social Sciences, Thousand Oaks, Califormia: Sage Publications, 93 p.

Allison, Paul D. (2009) Fixed Effects Regression Models, University of Pennsylvania, Sage Publications, Vol 160, 136 pp.

Border Environment Cooperation Commission and the North American Development Bank (1999) Status Report, Fall. El Paso, Tex: BECC, NADB.

Environmental Protection Agency. (1986) Ambient Water Quality Criteria for Bacteria. EPA 440/5-84-002.

Environmental Protection Agency. (2000). Revisions to Chapter 307 – Texas Surface Water Quality Standards.

Environmental Protection Agency. (2001). Status report on the Water-Wastewater Infrastructure Program for the US-Mexico Borderlands. EPA-832-R-00-001.

Environmental Protection Agency. (2002). Border 2012: U.S.-Mexico Environmental Program. EPA-160-D02-001.

Giner, Maria Elena (2009) "BECC: Analysis of Border Environmental Infrastructure Priorities, Now and to 2030" (keynote address at Border Institute X: Toward a Sustainable 2030).

Greene, William H. (2008) Econometric Analysis , Prentice Hall, 6th Edition, Upper Saddle River, N.J.

Hausman, Jerry A. 1978. "Specification tests in econometrics." *Econometrica* 46: 1251-1271

Hirsch, R.M., J.R. Slack, and R.A. Smith (1982), Techniques of Trend Analysis for Monthly Water Quality Data, Water Resources Research, 18(1) 107-121, DOI: 10.1029/WR018i001p00107

International Boundary and Water Commission (2008) Regional Assessment of Water Quality in the Rio Grande.

Mumme, S. "Innovation and Reform in Transboundary Resource Management: A Critical Look at the International Boundary and Water Commission, United States and Mexico," *Natural Resources Journal* 33, pp. 93-120, 1993

National Research Council (US) Committee on Indicators for Waterborne Pathogens. Indicators for Waterborne Pathogens. Washington (DC): National Academies Press (US); 2004. Appendix A, Emerging and Reemerging Waterborne Pathogens.

NORMA Oficial Mexicana (NOM). 1996. NOM-127-SSA1-1994: *Salud ambiental, agua para uso y consumo humano – Límites permisibles de calidad y tratamientos a que debe someterse el agua para su potabilización* (*'Environmental Health, water for human use and consumption – permissible limits of quality and treatments to which water must be submitted for its drinkability*'). Estados Unidos Mexicanos. Secretaría de Salud.

North American Agreement on Environmental Cooperation, U.S.-Canada-Mexico, Sept. 14, 1993, 32 I.L.M. 1480

North American Development Bank (1997a) "North American Development Bank Announces Cooperative Agreement with U.S. Environmental Protection Agency" (San Antonio, Texas: NADB)

North American Development Bank (1997b) *Protocol for Application of EPA Affordability Guidelines to U.S. Water and Wastewater Projects Under the Border Environment Infrastructure Fund* (San Antonio, Texas: NADB)

North American Development Bank (1997c) *Protocol for Application of EPA Affordability Guidelines to U.S. Water and Wastewater Projects Under the Border Environment Infrastructure Fund* (San Antonio, Texas: NADB)

North American Development Bank (2012) Annual Report.

North American Free Trade Agreement, U.S.-Canada-Mexico, September 14, 1993, 32 I.L.M. 1480 (1993)

Perez-Ciordia, I., Ferrero, M., Sanchez, E., Abadias, M., Martinez-Navarro, F., Herrera, P. (2002). *Salmonella* enteritis in Huesca; 1996-1999. [Article in Spanish] *Enfermedades infecciosas y microbiologia clinica*. 20(1): 16-21

Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley.

Schafer, J.L. (1997) *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall.

Stuart, S., McFeters, G., Schillinger, J. & Stuart, D. (1976). Aquatic Indicator Bacteria in the High Alpine Zone. *Applied and Environmental Microbiology* 31(2): 163-7

Texas Clean Rivers Program, United States Section International Boundary and Water Commission. (2003) Regional assessment of Water Quality in the Rio Grande Basin.

Texas Commission on Environmental Quality. Contract Number 582-2-44870/IBM-01-27

United States Department of Agriculture; United States Environmental Protection Agency (2014) *U.S. Mexico Border Needs Assessment and Support Project: Phase 1 Scoping Assessment Report*

United States Department of State (1993) Agreement Between the Government of the

United States Geological Survey, "National Hydrography Dataset," The National Map, U.S. Geological Survey, GIS Shapefile, http://viewer.nationalpmpa.gov/viewer/nhd.htm?p-nhd

United States of America and the Government of the Mexican States Concerning the Establishment of a Border Environment Cooperation Commission and a North American Development Bank November 16 and 18, 1993 WL 46884 (Treaty).

Varady, Robert (1996) "Post-NAFTA Opportunities and Challenges for the Environment: The First-Year of the Border Environment Cooperation Commission Part II," speech presented at the Tucson Committee on Foreign Relations.